

CERES Data Management Group Status Report

CERES Science Team Meeting

September 1st, 2015

Jonathan Gleason



Data Management Team

**Systems/
Optimization:**
Nelson Hillyer
Josh Wilkins
Mike Linsinbigler

**Production /
Optimization:**
Lisa Coleman
Carla Grune

Instrument:
Denise Cooper
Thomas Grepiotis
Dianne Snyder

Convolution:
Igor Antropov

Inversion:
Victor Sothcott

**Product Web
Support:**
Churngwei Chu
Ed Kizer
Cristian Mitrescu

**Configuration
Management:**
Tammy Ayers
Joanne Saunders

ERBE-like:
Dale Walikainen
Jeremie Lande

SARB:
Tom Caldwell

Task Manager:
Walt Miller

Clouds:
Sunny Sun-Mack
Ricky Brown
Yan Chen
Liz Heckert
Rita Smith
Sharon Gibson

TISA Gridding:
Raja Raju
Forrest Wrenn

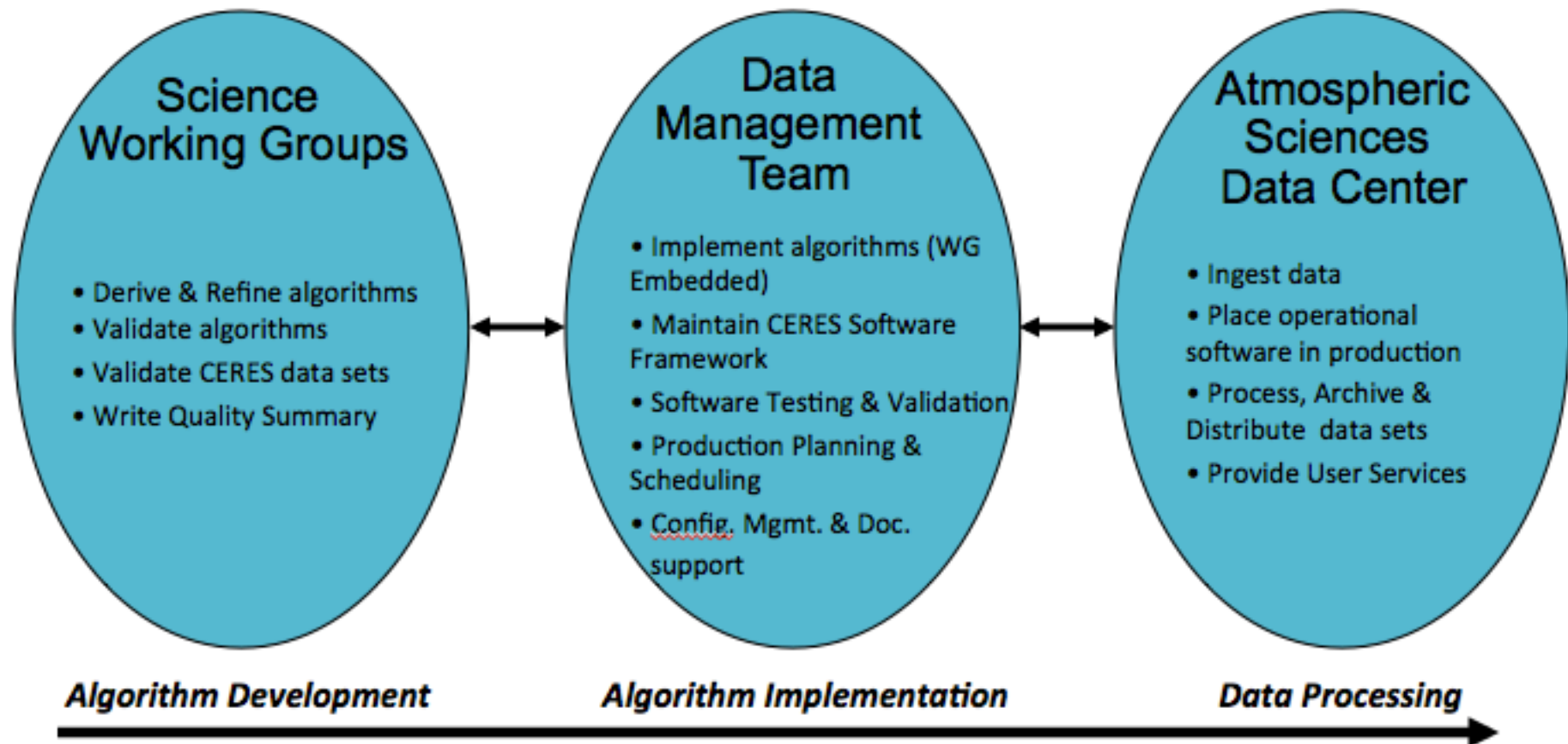
TISA Averaging:
Cathy Nguyen
Dennis Keyes



Overview

- Recent DMT Activities
- Current Data Availability
- Production Status Reporting
- Systems Status
- NPP Data Flow Changes to EDOS
- CATALYST v2.0 Status

Data Management Function



Currently ~1M production LOC contained in 54 active PGEs

CERES Production Software Overview

Subsystem Number	Subsystem Name	Number of PGEs	LOC (to nearest 1K)	Publicly Available Data Products	Product Frequency
1	Instrument/Pre-Processor	2	9K		
1	Instrument	6 + lib	162K	BDS	1/day
2	ERBE-like/ Inversion	5 + lib	31K	ES-8	1/day
3	ERBE-like/ TSA	2	14K	ES-9, ES-4	1/month
4.1 – 4.4	Clouds/VIIRS Subset Code	1	21K		12/hour
4.1 – 4.4	Clouds	11	358K		
4.5 – 4.6	Inversion	11	110K	SSF	1/hour
5	SARB	1	36K		1/hour
6 & 9	TISA-Gridding	7	34K	SSF1deg-Hour, ISCCP-D2like-Day/Nit	60/month, 36/month, 1/month
11	GGEO	1	6K	ISCCP-D2like-GEO	1/month
7.2	Synoptic SARB	1	47K		
7.1 & 8 10	TISA-Averaging	4	102K	SSF1deg-Day, SSF1deg-Month, SYN1deg-(3Hour, M3hour, Mhour, Month)	1/day, 1/month, 1/month, 5/month
12	MOA	2	14K		
	CERESlib/Perl_Lib		131K		
	Total	54	1,075K		



Recent Activity

22 Software & Data Deliveries Since May 5th, 2015

- **CATALYST (4)**
 - v2.0 Server, Ops Console & PGE Handlers
 - Intermediate Database updates
- **PR Tool (5)**
 - PR Tool v2.1 Web and DB deliveries
 - CATALYST v2.0 SSL support
 - Ed4 Variable display updates
- **Instrument (5)**
 - NPP Pre-Processor for EDOS Level-0 format
 - 2 scripting updates supporting EDOS NPP data flow testing
 - Postprocessor corrections for FM5
- **Clouds (1)**
 - Updates to Ed1 NPP PGEs
- **Inversion (2)**
 - Update Ed1 NPP PGE to run on dedicated P7 blades
- **TISA Gridding (2)**
 - Ed4 SSF1deg PCF corrections
 - Ancillary Xtk lookup tables
- **TISA Averaging (2)**
 - Ed3 SSF1deg redelivered to Use GMT and Ed4 format
 - Ed4 SSF1deg correction & x86 only
- **Perl_Lib (1)**
 - Support for new IBM P7 blades



Edition 3 Product Availability

Product	Platform	Processed through	Publicly Available
BDS	Terra & Aqua	May 31, 2015	Yes
SSF	Terra & Aqua	May 31, 2015 (FM1 & FM3 only)	Yes
SFC	Terra & Aqua	April 2015	Yes
SSF1deg-Day/Month	Terra & Aqua	June 2010	No*
SYN1deg	Merged	March 2015	Yes
ISCCP-D2like-Day/Nit	Terra & Aqua	April 2015	Yes
ISCCP-D2like-GEO	GEO	April 2015	Yes
ISCCP-D2like-Mrg	Terra + Aqua + GEO	April 2015	Yes
SSF Edition 1-CV	Terra Aqua	July 31, 2015 July 16, 2015	No



Edition 4 Product Availability

Product	Platform	Processed Through	Expected Current	Publicly Available*
BDS	Terra & Aqua	Dec 31, 2014	Nov 2015	Yes
SSF Beta2-Ed4	Terra Aqua	May 2, 2013 May 10, 2013	Dec 2015	No
SSF	Terra Aqua	Dec 31, 2012 Nov 30, 2012	Jan 2016	Yes
SSF1deg-Hour	Terra	Mar 2000 – Jul 2003 & Jan 2005 – Nov 2009	February 2016	No
	Aqua	July 2002 – Aug 2003 & Jan 2005 – Nov 2009		
SSF1deg-Day/ Month	Terra Aqua	Jan 2005 – Nov 2009	February 2016	Yes (as of 8/26)
SYN1deg	Merged	Jan – June 2008	June 2016	No
CldTypHist	Merged	Start January 2016	June 2016	

SYN1deg anticipated Release Date October 14th, 2015
CldTypeHist Pending Code Delivery



NPP Edition1 Product Availability

Product	Platform	Processed Through	Expected Current	Publically Available*
BDS	S-NPP	March 31, 2014	March 2016	Yes
SSF	S-NPP	March 31, 2012*	April 2016	Yes
SSF1deg-Hour	S-NPP	Start October 2015	May 2016	
SSF1deg-Day/Month	S-NPP	Start October 2015	May 2016	

* SSF Ed1 NPP processing was stopped in Spring 2015 to correct a problem that limited the number of footprints and processing resumed in August 2015



Production Monitoring



Production Status Reporting

CERES Monthly Processing objectives and status to be reported on Confluence

- Monthly page created by DMT outlining what to process by stream
- ASDC to provide weekly status updates each Monday

Plan to begin keeping 6-month processing objectives page on Confluence to address longer-term objectives

<https://asdcjira.larc.nasa.gov:9443/display/CARS/DMT+Processing+Objectives+-+August+2015>



System Status



System Status

ASDC System Maintenance Window September 18th – 23rd

Start Friday 9/18 @ 5:00pm - End Wednesday 9/23 @ 9:00am

- Upgrade all GPFS clients
- Firmware updates to IBM and SGI disk controllers
- Coil-blue hardware refresh
- Migrate IDL license server from asdsun to coil-blue

New IBM Flex blades available in AMI-P – currently “dedicated” for
Edition1 NPP SSF processing

- 6 IBM P7 Flex blades in production (2 in test area)
- 3 IBM x86 Flex blades in production (1 in test area)
- All Flex blades running Suse Linux 11.3 (vs 11.1)
- 20 additional x86 blades have been purchased for AMI-P



NPP-FM5 Data Flow Changes



NPP Data Flow Changes

Starting in September 2015, CERES FM5 data for NPP will no longer be obtained through NOAA IDPS

- EDOS to connect to JPSS ground station stream & package to EOS Level-0 format files
- All beta testing for CERES complete – Using small pre-processor PGE to merge 24-hour CERES file with 2-hour Attitude and Ephemeris files into 24-hour CERES files
- IDPS will no longer process RDRs

This approach anticipated to be used for FM6 on J1 and for RBI



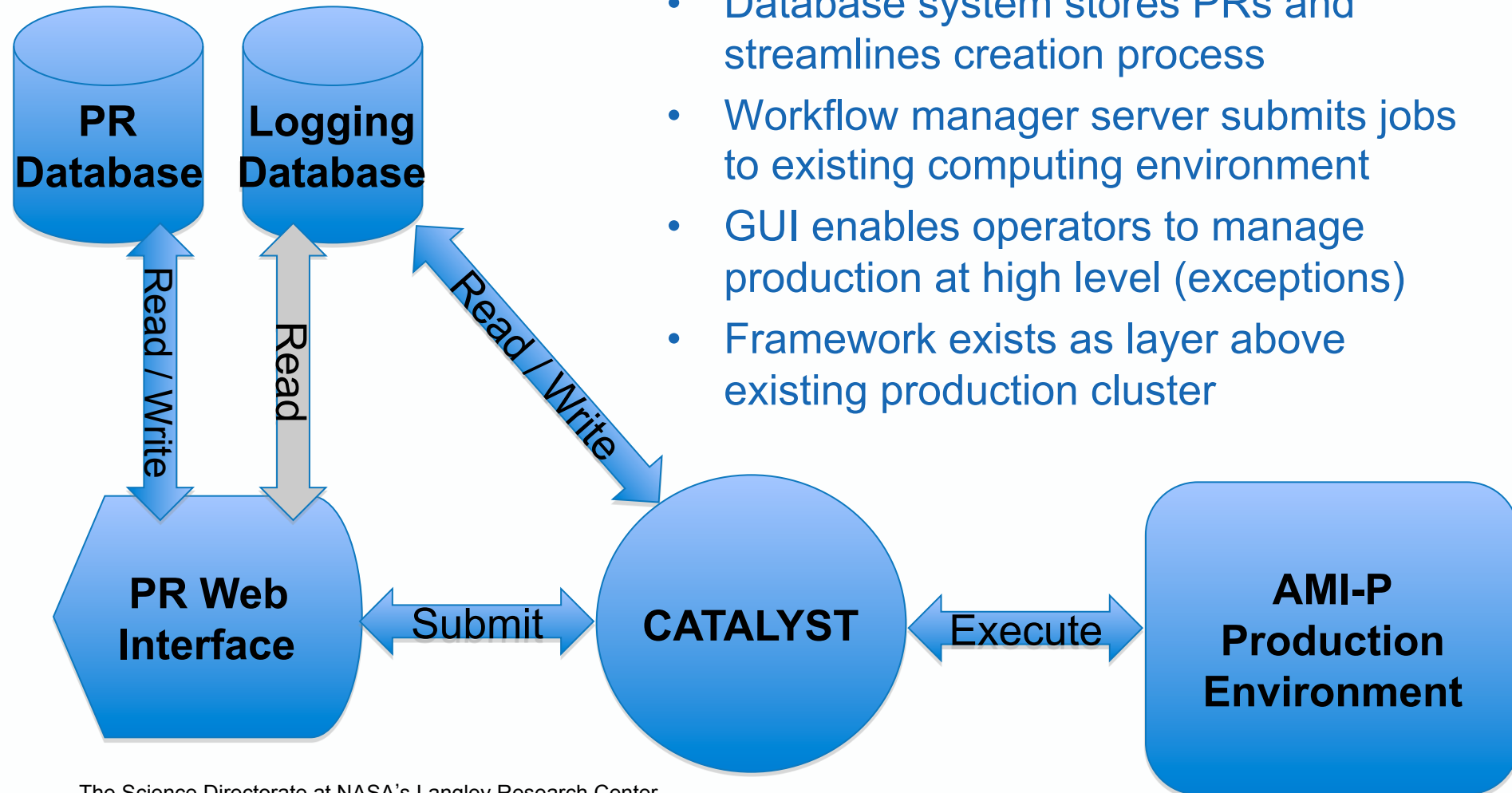
CATALYST Release 2.0



Automation Framework

CERES AuTomAted job Loading sYSTem (CATALYST)

- Database system stores PRs and streamlines creation process
- Workflow manager server submits jobs to existing computing environment
- GUI enables operators to manage production at high level (exceptions)
- Framework exists as layer above existing production cluster





CATALYST Status

Release 1.0 Design proof-of-concept Clouds & Inversion only

CATALYST Release 1.0 began running Beta2-Ed4 Clouds and Inversion April 17th, 2014

- 157 months Terra, 129 months Aqua
- >440,000 jobs successfully run

First build learning experience

- User feedback identified needed feature enhancements
- Bugs discovered

CATALYST Release 2.0 delivered for testing 8/24 and testing ongoing



CATALYST 2.0 Improvements

38 Change Requests submitted based on version 1.0

- 35 directly implemented or addressed in release 2.0
- 1 identified as too difficult to implement
- 2 delayed, likely to be superseded by release 2.0 capabilities

Improvements by category

- PR Validation
- Problem Isolation
- PGE Management
- Operator's Console Improvements
- Stability & Monitoring
- Logging
- Security



CATALYST 2.0 Improvements

PR Validation

- Extensive PR validation logic added – verify invalid or conflicting submitted PRs

Problem Isolation

Separate CATALYST to 6 sub-processes + Master process

- Limits client application downtime (i.e. PR Tool & Ops Console)
- Enables better logging and status reporting
- Improves Stability – Master processes monitors component health
- Contains known exception cases (i.e. server crashes, offline database, offline SGE)



CATALYST Build 2.0

Kernel – PR validation & coordination (PGE Handlers live here)

Cluster – Interfaces with SGE

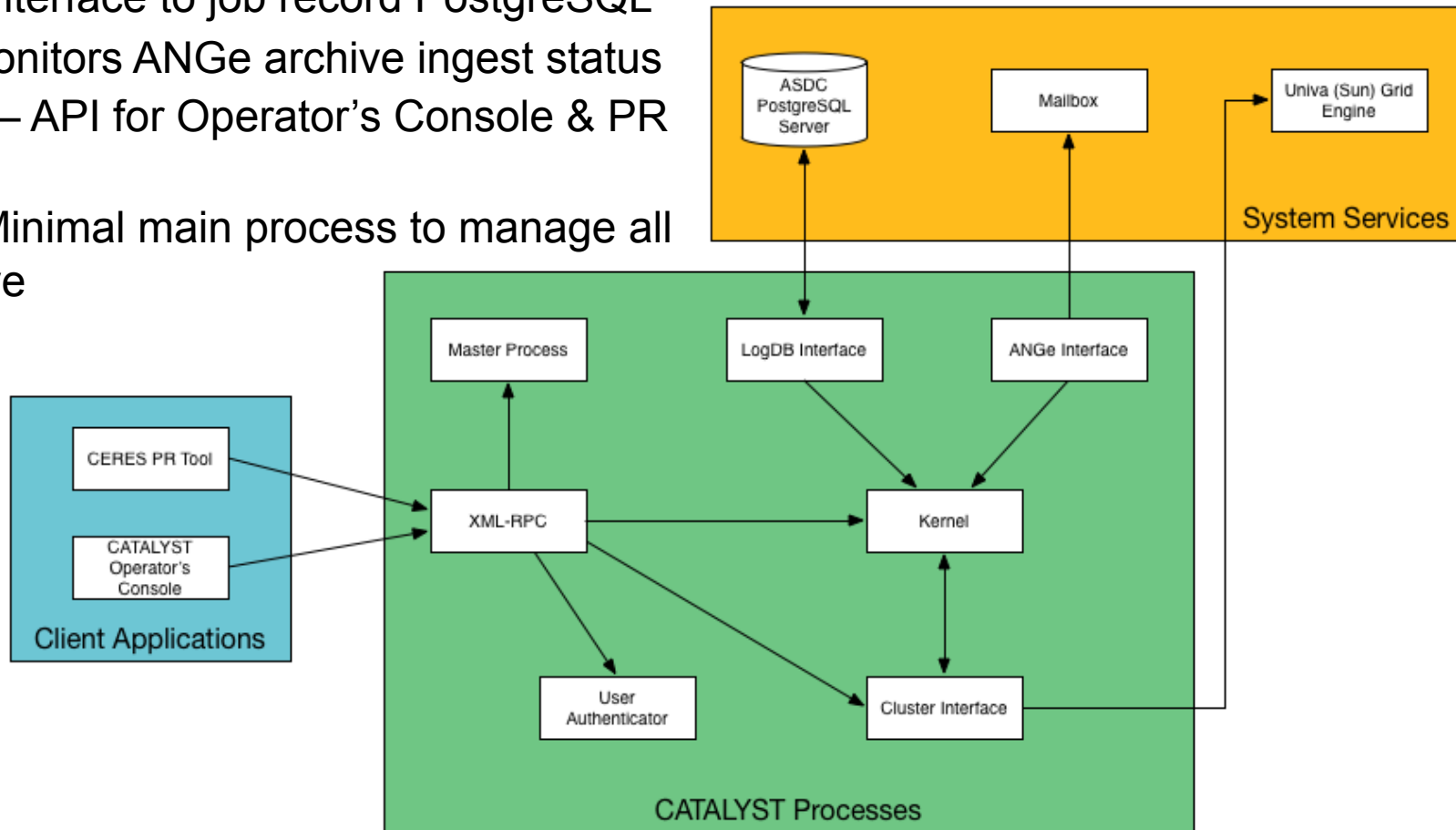
User – Authenticates users against LDAP

LogDB – Interface to job record PostgreSQL

ANGe – Monitors ANGe archive ingest status

XML-RPC – API for Operator's Console & PR Tool

Master – Minimal main process to manage all of the above

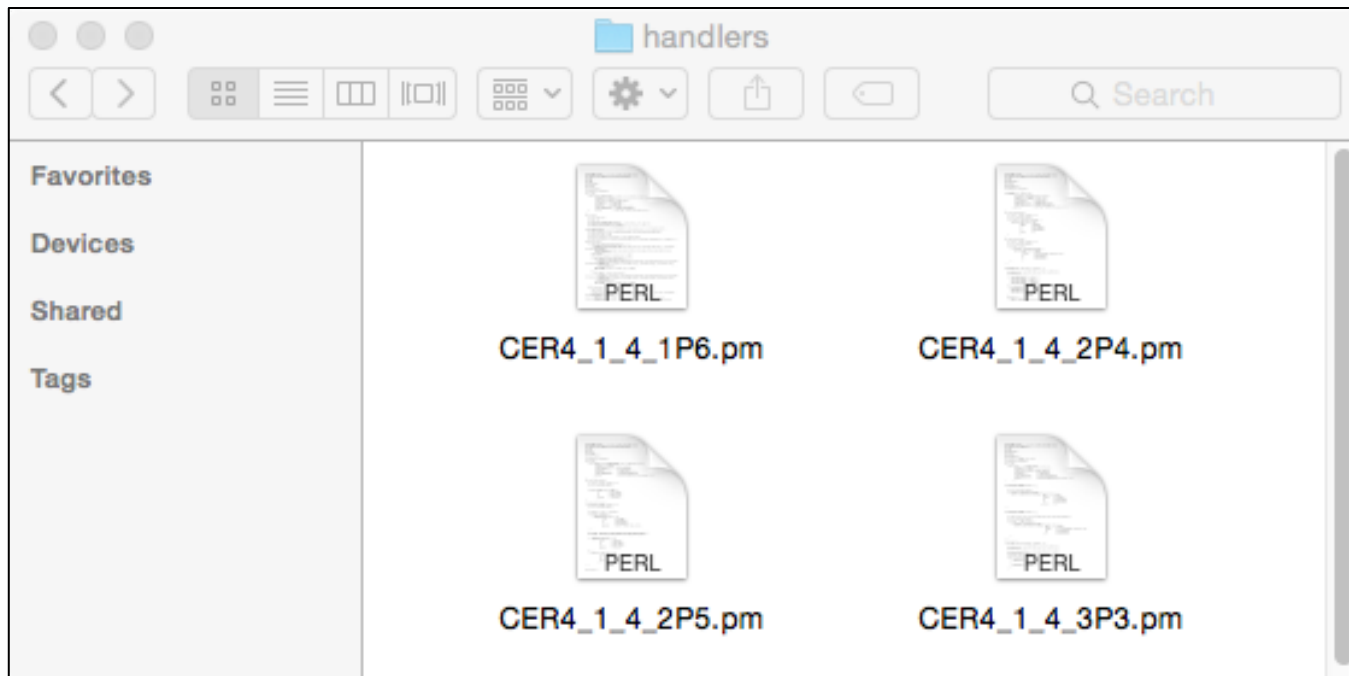




CATALYST 2.0 Improvements

PGE Management – file level

- Each PGE implemented as individual Perl module
- Modules installed in “CATALYST_HOME/handlers” directory
- PGE modules loaded on start of CATALYST kernel process
- Add/Remove PGE achieved by copy/delete .pm and restart kernel process





CATALYST 2.0 Improvements

PGE Management – Within Operator's Console

- Whole Subsystem or individual PGEs can be enabled or disabled within UI.
- Epilogue settings can be adjusted by subsystem or PGE
- PGE and Epilogue settings modifiable only by privileged CATALYST user (viewable by all)
- Epilogues run in FIFO order – if predecessor disabled follow-on will wait (prevents out of order file removal)

PGE Settings

Subsystem: Clouds

Subsystem Mode: ☒ Enabled ☐ Disabled

Subsystem Epilog Mode: ☐ Enabled ☒ Disabled

PGE: CER4.1-4.2P4

PGE Mode: ☒ Enabled ☐ Disabled

PGE Epilog Mode: ☐ Enabled ☒ Disabled

PGE Details:

available -> on
description -> Edition4 Daily QC Processor
granularity -> CATALYST::DataDate::Daily

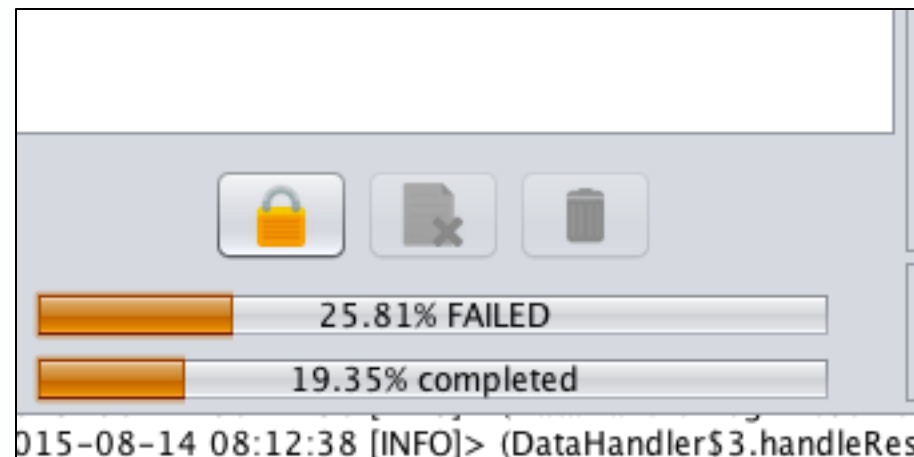
Apply Changes Close



CATALYST 2.0 Improvements

Operator's Console Improvements

- Added text search capability – search by date, job status, epilogue status or job ID
- Job status reporting
 - Added progress bars for jobs completed/failed
 - Color coded job listings (complete, in progress, failed, waiting)
- Progress bars and activity logs for user actions
- Read-Only Mode replaces need for v1.0 recovery utility – when server started, all job status' viewable before administrator initializes jobs
- Inline documentation with hovers text



Percentage failed/complete

CATALYST 2.0 Improvements

CATALYST Operator's Console (Connected - Server Uninitialized)

File CATALYST Server View Tools Help

Production Requests

- Non-Stream PRs
 - Clouds
 - CER4.1-4.2P4
 - CER4.1-4.2P4 Demo 01
 - CER4.1-4.3P3
 - CER4.1-4.3P3 Demo 01
 - Inversion
 - CER4.5-6.0P6 (PGE Disabled)
 - CER4.5-6.0P6 Terra-MODIS 14 (
 - CER4.5-6.2P3
 - CER4.5-6.2P3 Test PR FM3
 - CER4.5-6.4P2
 - CER4.5-6.4P2 Test PR FM2
 - SARB
 - CER7.2.1P2
 - CER7.2.1P2 Terra-Aqua-MODIS 1319-
 - CER7.2.1P2 Terra-MODIS 1318-15
 - TISA Averaging
 - CER7.3.1P2
 - CER7.3.1P2 Aqua-MODIS 1304-15
 - CER7.3.1P3
 - CER7.3.1P3 Terra-Aqua-MODIS 1308-
 - CER7.3.1P3 Terra-MODIS 1308-15
 - CER7.3.1P4
 - CER7.3.1P4 Terra-Aqua-MODIS 1313-
 - CER7.3.1P4 Terra-MODIS 1312-15
 - CER8.1P2
 - CER8.1P2 Terra-MODIS 1323-15

2005 -> 01 CER4.5-6.2P3 Test PR FM3 Last Refresh: 22:00:36

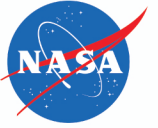
Instance	CATALYST Job Status	Science Status	Epilog Status	Deleted
20050101	completed	SUCCESS	SUCCESS	<input type="checkbox"/>
20050102	completed	SUCCESS	SUCCESS	<input type="checkbox"/>
20050103	completed	SUCCESS	SUCCESS	<input type="checkbox"/>
20050104	completed	SUCCESS	SUCCESS	<input type="checkbox"/>
20050105	failed_science	PGE_ERROR		<input type="checkbox"/>
20050106	failed_science	PGE_ERROR		<input type="checkbox"/>
20050107	failed_science	PGE_ERROR		<input type="checkbox"/>
20050108	failed_science	SIGSEGV		<input type="checkbox"/>
20050109	failed_science	PGE_ERROR		<input type="checkbox"/>
20050110	running			<input type="checkbox"/>
20050111	running			<input type="checkbox"/>
20050112	running			<input type="checkbox"/>
20050113	running			<input type="checkbox"/>
20050114	failed_epilog	SUCCESS	VERIFY_ERROR	<input type="checkbox"/>
20050115	running			<input type="checkbox"/>
20050116	waiting	SUCCESS	SUCCESS	<input type="checkbox"/>
20050117	waiting	SUCCESS	SUCCESS	<input type="checkbox"/>
20050118	waiting	SUCCESS	SUCCESS	<input type="checkbox"/>
20050119	waiting	SUCCESS	SUCCESS	<input type="checkbox"/>
20050120	epilog_running	SUCCESS		<input type="checkbox"/>
20050121	epilog_queued	SUCCESS		<input type="checkbox"/>
20050122	epilog_running	SUCCESS		<input type="checkbox"/>
20050123	waiting			<input type="checkbox"/>
20050124	waiting			<input type="checkbox"/>
20050125	waiting			<input type="checkbox"/>
20050126	waiting			<input type="checkbox"/>
20050127	queued			<input type="checkbox"/>
20050128	ready_to_launch			<input type="checkbox"/>
20050129	queued			<input type="checkbox"/>
20050130	waiting			<input type="checkbox"/>
20050131	waiting			<input type="checkbox"/>

This CATALYST job's science process has made an invalid memory reference.

16.13% FAILED

29.03% completed

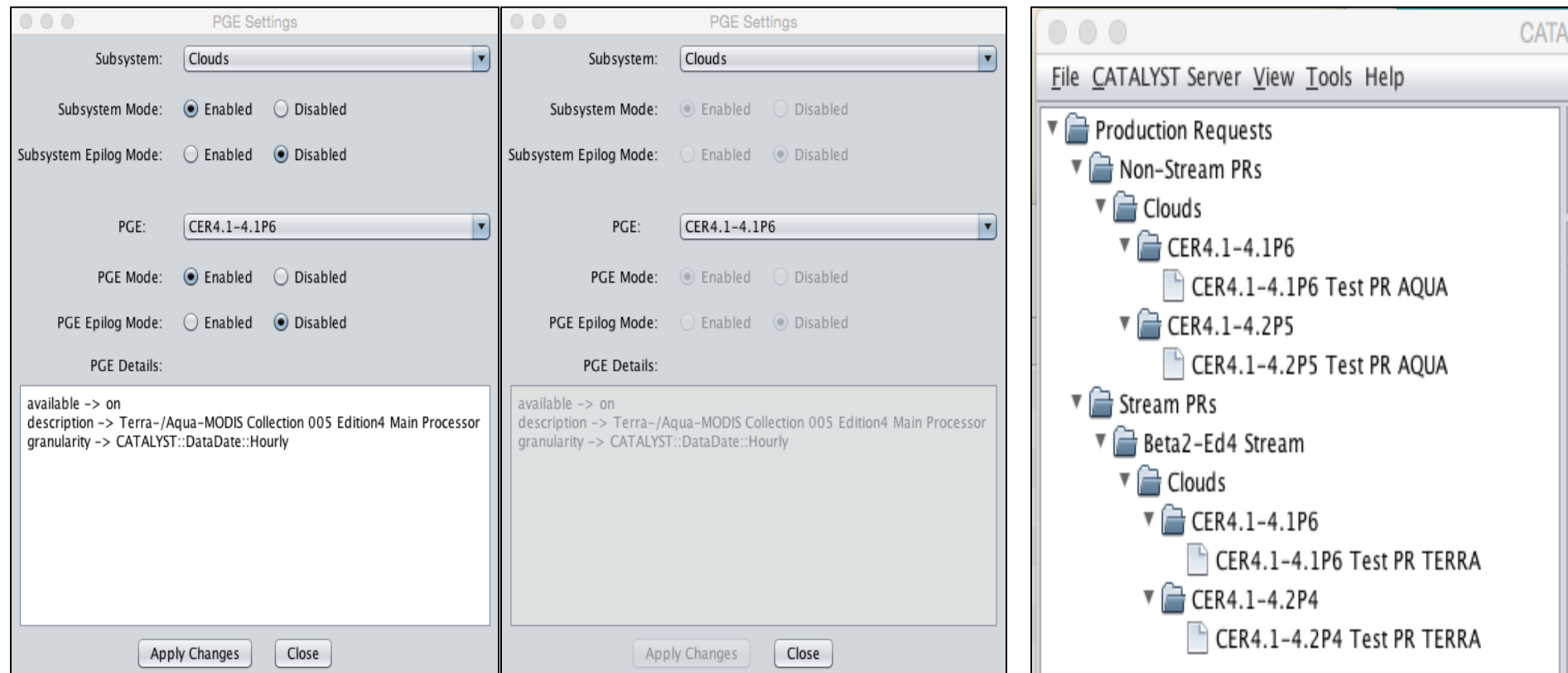
■ - Failed ■ - Success



CATALYST 2.0 Improvements

Operator's Console Improvements Cont'd

- PRs organized by processing stream
- Console window is resizable (previously static)
- Intelligent button disabling based on permissions



Functions grayed out if insufficient user permissions



CATALYST 2.0 Improvements

Stability & Monitoring

- Added auto-detection for inactive jobs in SGE grid engine
 - Detects if job manually deleted (“qdel-ed”) SGE jobs
 - Detects if blade issue causes job to hang
- New management utilities to monitor, start & stop CATALYST processes (Admin user only)
 - Graphical utility within Operator’s Console UI
 - Stand alone command line tool
- Improved system health monitoring email reporting
 - Generated by server process (previously standalone script)
 - More detailed and instructive info in reports – color coded to indicate status



CATALYST 2.0 Improvements

Logging

- Significantly improved logging capability
 - CATALYST Processes-level logging with all entries time stamped
 - Job details window now gives access within Console to:
 - PCF
 - PCF logs
 - LogStatus files
 - SGE logs
 - Epilogue logs

Security

- SSL support enabled for CATALYST API
 - All comm. From PR Tool & Operator's Console now encrypted end-to-end
 - No visible changes to user



CATALYST Schedule

Schedule Re-baselined late April 2015 (on schedule since)

TRR – August 17th

V2.0 Delivery – August 25th

Associated PR Tool Delivery – August 25th

Release to ASDC – September 4th

ASDC Testing – September 7th – 25th

ORR – October 5th

Build 2.0 Promote – October 13th

Deliver Ed4 SSF1deg & SYN1deg PGEs – October 23rd



Expanding CATALYST PGEs

PGE handlers to be delivered October 23rd 2015

- Edition 4 Inversion
- Edition 4 SSF1deg-Hr & Day/Month
- Edition 4 TSI, SARB & SYN1deg

PGE Handlers to be delivered December 2015

- Edition 1 NPP Clouds & Inversion

PGE Handlers to be delivered January 2016

- All Instrument and ERBE-Like PGEs



Backup Slides